



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

Ser Trp Phe Pro Val Gln Gly Pro Ala Asp Ile Cys Gln Cys Cys Asn Lys Gly Asp Gly Thr  
328 312 302

Pro Ser His Ser Arg Arg Glu Pro His Val Met Ser Gln Trp Ser Arg Ser Val Ser

338 348

### (57) Abstract

The invention relates to methods for producing recombinant human zona pellucida protein ("rhZP3") and glycosylated peptide having biological activity of binding to human spermatozoa. A human ovarian cell line is used to produce rhZP3 having a glycosylation pattern required for full biological activity. Methods of determining useful peptides with binding activity for human sperm and their syntheses, as well as using such peptides and proteins in therapeutics and diagnostics are discussed.

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## RECOMBINANT ACTIVE HUMAN ZONA PELLUCIDA PROTEIN 3 (hZP3)

Field of the Invention

This invention relates to male infertility testing, and to uses of recombinant human zona pellucida protein in clinical research and practical applications.

Description of Related Art

Fertilization is the process whereby individual gametes from the female (egg) and male (sperm) unite to create a zygote whose genetic makeup is different from both parents. The sperm-egg interaction requires zona pellucida protein 3 (ZP3) both as a sperm-oocyte binding ligand and as an acrosome reaction inducer.

In male patients with infertility of unknown etiology, an abnormal sperm-zona pellucida interaction between sperm and egg is observed frequently. This abnormality is associated with reduced sperm fertility capacity and may account for a significant proportion of infertility cases. Male infertility is a significant problem today, and approximately 30-40% of infertility cases can be attributed to male reproductive dysfunction. Therefore, there is a fundamental need to gain a deeper understanding of human sperm-oocyte interaction at the zona pellucida level. Greater understanding of this problem will offer improved and more physio-pathologically directed therapy to these patients.

Infertility in many cases arises from a problem with binding of egg with sperm to form a zygote. The binding of sperm to the zona pellucida of the egg is a crucial recognition event in this process that leads to fertilization. Extensive investigation of murine fertility systems has resulted in the identification and isolation of zona pellucida protein 3 ("ZP3") as the primary receptor for sperm within the zona pellucida Bleil and Wassarman, *Dev. Biol.* 76: 185-202 (1980). ZP3 is a glycopolypeptide which plays a crucial role during fertilization. As part of its biological role, ZP3 binds to a spermatozoon

and produces the acrosome reaction after less than 30 minutes of exposure to the spermatozoon. This biological activity consists of two parts. The first part is binding of ZP3 to a spermatozoon. The second part is induction of the acrosome reaction within a spermatozoon. The two parts can be detected by a number of procedures that are known to the skilled artisan.

Current evidence indicates that, as demonstrated in the murine model, ZP3 is involved in two events necessary for fertilization. First, ZP3 serves as the primary receptor for binding of sperm to the zona pellucida, Saling, *Oxf. Rev. Reprod. Biol.* 11: 339-388 (1989). Second, ZP3 is necessary for induction of the acrosomal reaction, Saling, *Biol. Rep.* 44: 246-251 (1991). Most studies in this area have used native ZP3 obtained from other animals. Reports on the human ZP3 glycopolypeptide indicate that this protein has a molecular weight of about 60 to 100 kD, with about half of this total as carbohydrate.

The biological role of ZP3 carbohydrate in fertility has been surmised but no definite conclusions pertaining to inter species discrimination have been made. In one recent experiment, mouse embryonal carcinoma cells were transfected with mZP3 in which five serine residues clustered in exon 7, Ser-329, Ser-331, Ser-332, Ser-333 and Ser-334 were converted to small non-hydroxy amino acids by site directed mutagenesis. The transgenic ZP3 was synthesized and secreted in an inactive form, having lost its ability to bind sperm, as described by Chen et al., *Proc. Natl. Acad. Sci. USA* 95: 6193-97 (1998). However, this work did not adequately address possible interspecies differences between ZP3s and did not adequately characterize differences between the five different serines. Moreover, the specificities of the UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase family which links the carbohydrate GalNAc to the side chain of certain serine and threonine residues in mucin type glycopolypeptides presently is unknown. However, empirical data already obtained with known mammalian peptide sequences can be used to predict the probability of glycosylation for a given sequence with high confidence, as reported by the Denmark Center for Biological Sequence Analysis. See for example, the center's website at [www.cbs.dtu.dk/services/NetOGlyc-2.0](http://www.cbs.dtu.dk/services/NetOGlyc-2.0), (hereinafter "NetOglyc website"). A database of O-glycosylated proteins can be found in Hansen *et al.*, *Nucl. Acids Res.*, 25: 278-282 (1997) and an algorithm for predicting glycosylation sites based on sequence context is found in Hansen *et al.*, *Glyco. J.*, 15: 115-130 (1998).

Advances in molecular biology have revealed new information about the ZP3 protein and its associated gene(s). Full-length cDNA clones of human ZP3 have been isolated, and the genomic loci of this gene has been characterized as described by Chamberlin and Dean, *Proc. Natl. Acad. Sci. U.S.A.* 87: 6014-6018 (1990). The genomic sequence from mouse has been obtained, and an exon 7 that codes the sequence at the carboxyl terminal end of the protein can be obtained by known methods. Kinloch *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 6409-13 (1988). Clones that contain sequence information from the mouse ZP3 gene express usable quantities of recombinant ZP3 ("rZP3") from, for example, the mouse, in tissue culture lines. Unfortunately, however, the production of recombinant human ZP3 ("rhZP3") is fraught with technical difficulties, not the least is the fact that expression of rhZP3 often leads to a generally unstable protein. Chapman and Barratt, *Mole. Human Repro.* 3:646, 648 (1997). Moreover, properly glycosylated hZP3 has never been purified before, so any information about this material is determined indirectly.

The apparent molecular weight of ZP3 can differ slightly depending on where it was made. For example, Beebe *et al.*, *Dev. Biol.* 151: 48-54 (1992), determined that recombinant murine ZP3 ("rmZP3") derived from transfected chinese hamster ovary ("CHO") cells was about 60 to 70 kDa, which differed in molecular weight from the average size of native ZP3, which is 83 kDa. These researchers attributed this incongruity to a difference in glycosylation pattern.

Despite the putative difference in glycosylation, rmZP3 demonstrated biological activity in homologous (mouse) spermzona pellucida competition assays, and rmZP3 was capable of triggering acrosomal exocytosis in capacitated mouse serum as described by Beebe *et al. Id.* When tested in a human system, however, rmZP3 displayed only a partial competitive effect under hemizona assay conditions, and did not function as an acrosome reaction agonist. Chamberlin and Dean, *Proc. Natl. Acad. Sci. USA* 87: 6014-18 (1990).

Another subsequent publication indicated the possible production of rhZP3 from CHO cells having biological activity. Van Duin *et al.*, *Biol. Repro.* 51: 607-17 (1994). However, the authors of the Van Duin report did not demonstrate that rhZP3 made has the capacity to bind sperm. Furthermore, the rhZP3 required at least three hours of contact with spermatozoa at an extremely high 10-20 µg/ml rhZP3 concentration before any significant acrosomal reaction could be seen, the ZP3 cannot be said to possess real

biological activity. This conclusion follows from the facts that biological activity is a specific (i.e. can occur at low concentration) binding of spermatozoa to ZP3 protein and also a quick acrosome reaction occurs. The specific binding and quick reaction are a necessary prelude to spermatocyte entry through the zona pellucida. The concentration of ZP3 under more natural biological conditions is at least 100 fold lower than reported by Van Duin, and the biological acrosomal reaction occurs in less than 30 minutes, not a minimum of a few hours.

One means of testing the biological reaction is with hemizona assay (HZA). HZA functionally tests for the assessment of tight binding of sperm to the zona pellucida, a critical step that triggers the physiological acrosome reaction leading to fertilization and early embryo development.

Because of the difficulties in obtaining suitable rhZP3 that could be used in such tests, most work has been carried out with genes from other species. Despite the problems it is now known that human ZP3, like mouse ZP3, comprises approximately 50% carbohydrate, (i.e. between 40% to 60% carbohydrate by weight) and the 424 amino acid long polypeptide differs from the mouse sequence by about one third. Also, it is generally understood that different host cell types may glycosylate hZP3 differently. However, "Whether such differences would alter the biological activity of rhZP3 is equivocal" as summarized recently. (Id. at p. 648).

Large quantities of biologically active hZP3 are needed for clinical applications that exploit the critical event of sperm binding to the zona pellucida, to diagnose a clinical condition. For example, a standardized and internally controlled "hemizona assay" tests this event in evaluating the binding capacity of human spermatozoa to human zona pellucida as described by Burkman *et al.*, *Fertil. Steril.* 49: 688-693 (1988) and Oehninger *et al.*, *Andrologia* 24: 307-321 (1992). Unfortunately, this need cannot be met from animal sources, or from presently available rhZP3. Thus, a procedure is desired to obtain large quantities of reproducible quality material with one or more biological properties of hZP3 both for research use and for diagnostic tests and pharmaceuticals that require a fully active protein.

SUMMARY OF THE INVENTION

It is an object of the present invention to provide glycopolypeptide having one or more biological properties of hZP3. It is another object to provide isolated and biologically functional rhZP3 protein. It is yet another object to provide rhZP3 glycopolypeptide that binds human sperm and that has diagnostic and therapeutic use related to human fertility. Yet another purpose is to provide methods for testing male infertility, based on induction of the acrosome reaction caused by binding of biologically active hZP3 to clinical sample materials. Yet another purpose is to provide high levels of biologically active hZP3 of consistent quality for use in the clinical and research chemistry.

In achieving these objectives and other objects, the invention provides a recombinant glycopolypeptide of about 65 kd to about 100 kd that comprises approximately 50% carbohydrate in terms of weight and that can bind human spermatozoa and induce an acrosome reaction. The invention also provides a recombinant glycopolypeptide of about 65 kd to about 100 kd that comprises approximately 50% carbohydrate by weight and that can bind human spermatozoa and induce an acrosome reaction, wherein the glycopolypeptide pattern is produced by a cell having glycosylation machinery similar to that of an oocyte. The invention further provides a process for producing a glycopolypeptide having the biological activity of hZP3 protein, comprising the steps of: transfecting an ovarian cell line cell with a gene that encodes a hZP3 polypeptide; culturing the cell to produce a culture of hZP3 producing cells; and isolating the glycopolypeptide.

The invention also includes a method of detecting male infertility from a semen sample, comprising the steps of: contacting a solution of recombinant glycopolypeptide of about 65 kd to about 100 kd that comprises approximately 50% carbohydrate by weight and that can bind to human spermatozoa and induce an acrosome reaction with spermatozoa from the sample to form an admixture; detecting the acrosome reaction and acrosomal lysis in the admixture; and comparing the amount of acrosomal lysis with a reference value. The invention further provides a cell from a transformed human ovarian cell line comprising a rhZP3 gene.

Another embodiment of the invention is a transgenic glycopolypeptide having an active portion (i.e. glycosylated amino acid sequence region) that preferentially binds to human sperm, wherein the active portion has a polypeptide size of less than 25kDa, comprises an amino acid sequence that is more than 54% identical with SEQ ID NO: 1 and

exhibits a predicted O-glycosylation site at the fifth position from the carboxyl terminus. Yet another embodiment is a method of detecting male infertility from a spermatozoa sample, comprising the steps of: (a) contacting transgenic glycopolypeptide having an active portion that preferentially binds to human sperm, wherein the active portion has a polypeptide size of less than 25kDa, comprises an amino acid sequence that is more than 54% homologous with SEQ ID NO: 1 and exhibits a predicted O-glycosylation site at the fifth position from the carboxyl terminus, with spermatozoa from said sample to form an admixture; (b) detecting biological activity of said spermatozoa in said admixture; and (c) comparing the biological activity with a reference value. Yet another embodiment is a nucleic acid vector useful for producing a glycopolypeptide, the glycopolypeptide specifically binds to human sperm, wherein the vector codes for an amino acid sequence that is more than 54% homologous with SEQ ID NO: 1. Yet another embodiment is a human cell containing a nucleic acid vector useful for producing a glycopolypeptide less than 200 amino acids long, wherein the glycopolypeptide specifically binds to human sperm and the vector codes for an amino acid sequence that is more than 75% identical with SEQ ID NO: 1.

#### DESCRIPTION OF THE FIGURES

Figure 1 is the amino acid sequence for residues 308 to 348 of human ZP3.

Figure 2 is a representative prediction summary of O-glycosylation sites on the 308-348 amino acid region of human ZP3.

Figure 3 is a representative prediction summary of O-glycosylation sites on the 309-349 amino acid region of mouse ZP3.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The inventors have discovered that tissue-specific and species-specific differences in glycosylation pattern of the hZP3 protein, and more particularly the role of residues 308 through 349, greatly affect the biological activity of this protein. From this discovery, the inventors have developed a cell expression system that uses the ovarian glycosylation machinery to properly glycosylate hZP3. The inventors also have developed glycopolypeptides that bind human sperm. "Properly glycosylate," in this context means to

give a glycosylation pattern that is similar to that of the human oocyte (is "human-functional") such that the glycopolypeptide has biological activity (i.e. specific binding to oocyte) with human sperm. Using the present system, the inventors have isolated rhZP3 for the first time that, unlike previous rhZP3, is fully biologically active and contains carbohydrate that more closely resembles human oocyte protein compared to previously known rhZP3. The inventors also have developed a test using the biologically active recombinant glycopolypeptide to diagnose causes of male infertility. Furthermore, the inventors have discovered therapeutic uses of rhZP3 and rhZP3 glycopolypeptides smaller than rhZP3 that have not been realized before.

Based on their understanding of the role of hZP3 residues 308 through 349 in specifying human oocyte glycosylation, the inventors have discovered alternative biologically functional glycopolypeptides that simulate binding of human oocyte to human sperm. The term "biologically functional glycopolypeptide" in this context means a polypeptide that comprises at least the segment of amino acids having the sequence SEQ ID NO: 1 and which binds to human sperm better than to mouse sperm. Preferably the polypeptide further comprises carbohydrate that has been added by a human cell during polypeptide synthesis. More preferably the human cell is from an ovary or follicle cell line. In one embodiment the cell line is a non-ovarian mammalian (and preferably human) cell line that has been genetically altered for the induction and/or stimulation of oocyte glycosylation enzymes. In this context, large scale production of small glycopolypeptides of less than 200 amino acids having sequence identity of more than 54% and preferably more than 75% with SEQ ID NO:1 allows new uses such as, for example, contraception, whereby the glycopolypeptide interferes with normal fertilization.

The inventors discovered that the biological specificity of rhZP3 predominantly comes from the carbohydrate portion of this section of hZP3. More specifically, the inventors have discovered that human rZP3 made by non-human ovarian cell lines such as a CHO cell line, while having the same amino acid sequence as human ZP3, are not active with human eggs because of their carbohydrate component. For example, removal of carbohydrate from hZP3 by a glycosidase will remove the biological activity of the hZP3. The precise differences in glycosylation can be determined by first making a hZP3 having full biological activity. Then, the carbohydrate portion of the glycopolypeptide can be

partly altered or removed. After alteration, the glycopolypeptide is tested to see if it retains full activity with human eggs.

The inventors' knowledge of the subtle biochemical differences between human and mouse ZP3 that directly account for the interspecies discrimination of this protein, provides new tools and techniques for both creating and using glycosylated glycopolypeptide for diagnoses and therapies related to human fertility. More specifically, the primary sequence of ZP3, contrary to many held beliefs, directs a different glycosylation, which, together with the related and unique species-specific and tissue-specific glycosylation machinery in the human oocyte, provides a unique human oocyte glycosylation pattern. Thus, one aspect of the invention is the discovery that human specific oocyte glycosylation in human oocytes does indeed matter when expressing the ZP3 protein. Another aspect of the invention is that the amino acid sequence within residues 308 to 348 of ZP3 effects and directs human specific oocyte glycosylation, being species specific. Yet another aspect of the invention is that changes to this specific sequence can be made according to predictive features of an algorithm specifically designed to look for the effect of sequence on the probability of O-glycosylation within that sequence.

The discovery of a species-specific binding region within hZP3 provides the ability to improve potency of glycopolypeptides that mimic the one or more biological activities of hZP3. Thus, another aspect of the invention is that a smaller glycopolypeptide, comprising less than 400 amino acids can be expressed having the binding ability of human ZP3 by virtue of the presence of an active region that is glycosylated in a human-functional pattern (i.e. with human-like carbohydrate suitable for species specific function with human spermatozoa). Specific human-like carbohydrate useful in this context is characterized and explained by, for example, Clark, et al., Human Reproduction 11: 467-473 (1996); Clark et al., Molecular Human Reproduction 2: 513-517 (1996); Patankar et al., Molecular Human Reproduction 3: 501-505 (1997); and Ozgur et al., Molecular Human Reproduction 4(4): 318-324 (1998).

In advantageous embodiments the glycopolypeptide is between 41 to 400 amino acids long. When used primarily for the species-specific binding reaction, the glycopolypeptide should be smaller, such as between 41 to 300, 50 to 200, 50 to 150 and even 50 to 100 amino acids long to allow greater binding sites for a given amount of material. Most advantageous for binding reactions is a glycopolypeptide that shares at least

54% sequence identity, particularly greater than 75% sequence identity, and most particularly greater than 80% sequence identity with SEQ ID NO: 1 and that is between 41 and 65 amino acids long.

The glycosylation advantageously is carried out by expression in human cells (i.e. cells that express human glycosylation enzymes) and particularly is carried out by expression in a human follicle cell line or ovarian cell line such as PA-1. In other embodiments, the glycopolypeptide that contains this region, shown in SEQ ID NO:1, can be as small as 41 amino acids long. However, advantageously, a large glycopolypeptide up to 200, 400 or even more amino acids long or may be made from repeated units of the binding portion shown in SEQ ID NO: 1.

In the context of using the species-specific region, the inventors discovered 3 parameters. One, residues 308 to 348 of ZP3 contain interspecies information and nucleic acid vectors that contain the human residues 308 to 348 of ZP3 are particularly useful to make polypeptides having human-specific glycosylation. Two, the amino acids encoded by residues 308 to 348 as shown in SEQ ID NO: 1, differ from that of ZP3 found in other species, particularly in the capability of serines to become O-glycosylated by cellular machinery. Three, in contrast to N-glycosylation, the pattern of predicted O-glycosylation of the residues 308 to 348 carboxyl terminal serines in peptides or peptides comprising the sequence greatly affects the actual glycosylation of a polypeptide.

The inventors discovered that the human ZP3 sequence has a serine 344 residue which is most likely glycosylated. In contrast, a corresponding sequence from another species such as, for example, the mouse, shows a greatly different predicted O-glycosylation probability. For example, the analogous residue (345 serine) in mouse has drastically less predicted glycosylation, and instead, the 332 serine stands out as being most likely glycosylated. Thus, the discovered technique can be used to distinguish and to predict which sequences and sequence modifications can provide a suitable human sequence that will become glycosylated in human cells, and in particularly cells from a human ovary or follicle cell line. Figures 2 and 3 depict the respective predictive O-glycosylation parameters. As seen in these figures, the higher the potential value with respect to its corresponding threshold value, the greater chance that the amino acid will react with an N-acetylgalactosaminyltransferase. More specifically, the "potential" value is a relative

measure of whether the designated residue should be O-glycosylated. The "threshold" value is a relative measure that inversely relates to the chances of glycosylation.

The inventors point out that the output shown in Figures 2 and 3 reflects empirical knowledge obtained from proteins found in other, non-ovarian cells and is a relative measure. Clearly this analysis shows that, based on the influence of sequence on glycosylation in other systems, the glycosylation of residues 308 to 348 of human ZP3 is primarily directed to position 344, whereas the corresponding mouse sequence has glycosylation directed to 13 positions to the amino terminal side of that serine residue. Furthermore, altering the human sequence in a manner to remove this calculated propensity for O-glycosylation will also decrease the affinity of the glycosylated polypeptide for the human sperm surface. The analysis used here can be used to derive other sequences contemplated as the invention. In particular, amino acid substitutions can be made to residues at positions listed in Table 1, which preserve the unique human-species specific glycosylation of the glycopolypeptide of the invention. This table lists representative amino acids that, according to the algorithm can be substituted while maintaining the unique human glycosylation pattern.

For most positions, exemplified by position 326, a conservative amino acid substitution can be made. By "conservative" substitutions is meant replacing an amino acid residue with another that is biologically and/or chemically similar, e.g., one hydrophobic residue for another, or one polar residue for another. The substitutions include combinations such as Gly, Ala; Val, Ile, Leu; Asp, Glu; Asn, Gln; Ser, Thr; Lys, Arg; and Phe, Tyr. Preferably, the portion of the sequence that is intended to mimic substantially the peptide of SEQ ID NO: 1 will differ less than 50% from SEQ ID NO:1, except where additional amino acids may be added at either terminus.

A considerable amount of work in this area has provided algorithms to use in making such changes. For example, the hydropathic index of amino acids may be considered. The importance of the hydropathic amino acid index in conferring interactive biologic function on a protein is generally understood in the art as cited in U.S. No. 5,703,057. It is accepted that the relative hydropathic character of the amino acid contributes to the secondary structure of the resultant peptide which in turn defines the interaction of the peptide with other molecules, for example, receptors, DNA, glycosidases, and the like. Each amino acid has been assigned a hydropathic index on the basis of its

hydrophobicity and charge characteristics, these are: isoleucine (+4.5); valine (+4.2); leucine (+3.8); phenylalanine (+2.8); cysteine/cystine (+2.5); methionine (+1.9); alanine (+1.8); glycine (-0.4); threonine (-0.7); serine (-0.8); tryptophan (-0.9); tyrosine (-1.3); proline (-1.6); histidine (-3.2); glutamate (-3.5); glutamine (-3.5); aspartate (-3.5); asparagine (-3.5); lysine (-3.9); and arginine (-4.5).

It is known in the art that certain amino acids may be substituted by other amino acids having a similar hydropathic index or score and still result in a peptide with similar biological activity, i.e., still obtain a biological functionally equivalent peptide. In making such changes, the substitution of amino acids whose hydropathic indices are within +- 2 is preferred, those which are within +- 1 are particularly preferred, and those within +- 0.5 are even more particularly preferred. It is also understood in the art that the substitution of like amino acids can be made effectively on the basis of hydrophilicity. U.S. Pat. No. 4,554,101, for example, states that the greatest local average hydrophilicity of a protein, as governed by the hydrophilicity of its adjacent amino acids, correlates with a biological property of the protein. As detailed in U.S. Pat. No. 4,554,101, the following hydrophilicity values have been assigned to amino acid residues: arginine (+3.0); lysine (+3.0); aspartate (+3.0 +- 1); glutamate (+3.0 +- 1); serine (+0.3); asparagine (+0.2); glutamine (+0.2); glycine (0); threonine (-0.4); proline (-0.5 +- 1); alanine (-0.5); histidine (-0.5); cysteine (-1.0); methionine (-1.3); valine (-1.5); leucine (-1.8); isoleucine (-1.8); tyrosine (-2.3); phenylalanine (-2.5); tryptophan (-3.4).

It is understood that an amino acid can be substituted for another having a similar hydrophilicity value and still obtain a biologically equivalent peptide. In such changes, the substitution of amino acids whose hydrophilicity values are within +- 2 is preferred, those which are within +- 1 are particularly preferred, and those within +- 0.5 are even more particularly preferred. As outlined above, amino acid substitutions are generally therefore based on the relative similarity of the amino acid side-chain substituents, for example, their hydrophobicity, hydrophilicity, charge, size, and the like. Exemplary substitutions which take various of the foregoing characteristics into consideration are well known to those of skill in the art and include: arginine and lysine; glutamate and aspartate; serine and threonine; glutamine and asparagine; and valine, leucine and isoleucine. In the present case, the inventors discovered that, with the exception of substitutions shown in Table 1, serines in SEQ ID NO: 1 cannot be substituted because of their role in glycosylation.

A "biologically functional" glycopolypeptide or rhZP3, in the context of the present invention means that the polypeptide (containing as a minimum amino acids 308 to 348, and functional derivatives thereof) or the ZP3 (if the whole protein is made) binds spermatozoa when present at a concentration below 1  $\mu\text{g}/\text{ml}$ , and induces an acrosome reaction within about one hour upon binding. Advantageously, the rhZP3 binds when present at a concentration below 0.1  $\mu\text{g}/\text{ml}$  and most advantageously when present at a concentration of 0.01  $\mu\text{g}/\text{ml}$  to 0.1  $\mu\text{g}/\text{ml}$ . The time of the acrosome reaction can be less than 1 hour or even less than 30 minutes. This binding of ZP3 protein to spermatozoa is species specific and also can be used to detect deficiencies in spermatozoa that account for male infertility.

An "ovarian cell line" as termed here means a human cell line from an ovary tissue such that cells of this line produce glycopolypeptide having characteristic carbohydrate of oocyte glycopolypeptides. An oocyte-like glycopolypeptide carbohydrate pattern can be compared with a non-oocyte-like glycopolypeptide pattern by any of a number of techniques well-known to a skilled artisan.

Methods for comparing glycopolypeptide patterns in the context used above include, for example, the use of monoclonal antibodies, as described by Barbosa to detect changes in the carbohydrate portion of HLA glycopolypeptide, using oligonucleotide-directed mutagenesis, gene transfer techniques (Barbosa *et al.*, *J. Exp. Med.* 166: 1329-50 (1987)) and newer biochemical methods to directly assess carbohydrate content.

Another method that can be used to directly compare whether a recombinant glycopolypeptide has a correct glycosylation pattern is to determine whether the glycopolypeptide can block sperm binding to native oocytes in a "hemizona assay." In normal conditions without the glycopolypeptide, the sperm binds to oocytes. However, when glycopolypeptide having proper oocyte glycosylation is incubated with the sperm, the glycopolypeptide will bind to the sperm first and block the sperm from further binding or reacting with the oocyte. In contrast, if the glycopolypeptide lacks a correct glycosylation pattern, the glycopolypeptide will not block sperm binding to the oocyte.

Yet another method of determining whether a glycopolypeptide possesses a correct glycosylation pattern is to remove carbohydrate side chains from the glycopolypeptide with peptide N-glycosylase F and O-glycanase and then incubate reaction products with sperm. If the glycopolypeptide preparation can block sperm binding to a human oocyte before

digestion but cannot after digestion, then the undigested preparation possesses a correct glycosylation pattern.

A "recombinant human zona pellucida protein 3" means a transgenic protein (or native protein under the control of a transgenic control element) that comprises at least part of the hZP3 gene and which retains substantially all of the hZP3 glycopolypeptide activity. "Substantially all" means that the protein exhibits binding activity and stimulation of the acrosome reaction when co-present with spermatozoa in aqueous solution at a concentration of less than about 1 µg/ml (e.g. 1 µg/ml), and particularly less than about 0.1 µg/ml (e.g. 0.1 µg/ml), for a time period of less than about 1 hour (e.g. 1 hour), and particularly, less than about 30 minutes (e.g. 30 minutes).

DNA sequences useful for complete hZP3 transgenic protein expression are known by the skilled artisan and derived from the known 424 amino acid sequence of hZP3, which is described by Chamberlin and Dean, *Proc. Nat. Acad. Sci. U.S.A.* 87: 6014-6018 (1990). The transduction of cells from an ovarian cell line with such a sequence can be effected via known methods. Generally, a DNA construct is used that contains a promoter upstream of the structural gene that encodes the desired protein sequence. Suitable promoters are described, for example, in U.S. patent No. 5,618,698. According to this embodiment, both a rhZP3 structural gene and a regulatory element to control the gene are transduced into the host cell.

According to another embodiment of the invention, a cell from an ovarian cell line is transduced *ex vivo* with a DNA comprising a promoter and a homologous DNA (but not an intact structural gene) that can link up with and function (i.e., turn on or increase expression) with an endogenous gene within the nucleus of the cell. In this embodiment, DNA comprising a regulatory sequence, an exon and a splice donor are introduced into a cell by homologous recombination into the cell's genome at a preselected site. The introduction of this DNA results in the production of a new transcription unit in which the regulatory sequence, exon and splice donor site are operatively linked to the endogenous gene.

The phrase "operably linked" refers to a first sequence(s) being positioned sufficiently proximal to a second sequence(s) so that the first sequence(s) can exert influence over the second sequence(s) or a region under control of that second sequence. For instance, a regulatory sequence can be operably linked to a promoter, whereby this

sequence enhances the transcriptional strength of the promoter. In this situation, the regulatory sequence would typically be 5' to the promoter. The regulatory sequence and promoter can, in turn, be operably linked to a gene so that the gene will be expressed under the control of the regulatory sequence/promoter combination, which would typically be 5' to the gene.

The introduction of DNA typically is followed by selection of cells that have received a promoter in a desired location to turn on the desired gene. Applicable selection methodology is described, for instance, in U.S. patent Nos. 5,641,670 and 5,272,071. Selection techniques also are described by Mansour *et al.*, *Nature* 136: 348, 349 (1988). After selection, the cells which express the desired gene are cultured and the expressed gene product is harvested.

According to preferred embodiments the DNA sequence useful for hZP3 transgenic expression codes for a shorter protein (polypeptide) molecule, (i.e. less than 25kDa polypeptide, particularly less than 10kDa polypeptide and more particularly less than 5kDa). Such preferred protein sequences include a core region as shown in SEQ ID NO: 1. The inventors realized that SEQ ID NO: 1 is the most important determinant with respect to the inter-species glycosylation pattern of ZP3 and that a small protein having this sequence, for a given mass, will be particularly potent in binding sperm. The SEQ ID NO:1 is 54% homologous with the corresponding sequence from mouse ZP3. By "homologous" is meant that when the two sequences are compared, 54% of the amino acids are the same. Despite the homology, the mouse ZP3 has a small amount of binding with human sperm, although this affinity is less than one-tenth, and typically less than one-hundredth the level observed with human ZP3. The term "can strongly bind human spermatozoa" is used herein to mean binding that is at least about 10 times (e.g. 10 times) as strong as an equivalent molar amount of mouse ZP3. In practice, such a qualitative determination is carried out by incubating different concentrations of material in a binding assay and determining binding directly or indirectly by competition. Such binding assays are exemplified in the examples herein and are well known to the skilled artisan. Thus, a peptide sequence that differs from human ZP3 by up to 46% maintains some residual human sperm binding activity that is stronger than the equivalent molar amount of mouse ZP3. Accordingly, a peptide according to one embodiment is more than 46% identical to the human ZP3 sequence shown in SEQ ID NO: 1. Furthermore, it is preferred that any

deviation from this sequence be limited to positions 3, 8, 13, 16, 17, 19, 21-23, 25, 27, 28, 30, 32-35 38 and 39, (positions listed in Figure 1) as the non-listed positions are more conserved.

In another embodiment the sequence is identical with the last 11 amino acids at the carboxyl terminal end of the portion depicted in Figure 1. This is because the serine in the middle of this sequence is particularly important for the human specific glycosylation.

In yet another embodiment the invention is a nucleic acid reagent vector that comprises a DNA sequence that codes for SEQ ID NO: 1. Such vector advantageously is a carrier for delivery of a gene sequence for expression of the ZP3 binding activity. In one embodiment the vector comprises residues 308 to 348 but lacks the remainder of the ZP3 coding gene. Construction of such a carrier is known to the skilled artisan as described, for example in the references cited herein. According to this last embodiment a glycosylated polypeptide useful for diagnostics and pharmaceutics related to effects on binding sperm and human fertility, is contemplated that is made by a vector comprising the human residues 308 to 348 or other DNA that codes for SEQ ID NO: 1.

The term "rhZP3 analog" refers to hZP3 mutants and chemically altered derivatives that have the above-listed biochemical and biological attributes of native rhZP3. In particular, changes in the amino acid sequence of hZP3 are contemplated in the present invention. hZP3 can be altered by changing the DNA encoding the protein. Preferably, only conservative amino acid alterations are undertaken, using amino acids that have the same or similar properties and the glycosylation recognition sites are not altered as described above.

Additionally, other variants and derivatized species of hZP3 can be used in the present invention. Variants include analogs, homologs, derivatives, muteins and mimetics of hZP3 that retain the ability to bind spermatozoa and induce the acrosome reaction. Fragments of the hZP3 refer to portions of the amino acid sequence of hZP3 that also retain this ability. The variants and fragments can be generated directly from hZP3 itself by chemical modification, by proteolytic enzyme digestion, or by combinations thereof. Additionally, genetic engineering of non-ovarian cell lines to alter their protein glycosylation machinery to emulate that of the oocyte could be carried out to allow expression of hZP3 protein in non-oocyte cells. '

Variants and fragments (i.e. analogs) of rhZP3 also can be created by recombinant techniques employing genomic or cDNA cloning methods. Site-specific and region-directed mutagenesis techniques can be employed. See CURRENT PROTOCOLS IN MOLECULAR BIOLOGY vol. 1, ch. 8 (Ausubel *et al.* Eds., J. Wiley & Sons 1989 & Supp. 1990-93); PROTEIN ENGINEERING (Oxender & Fox Eds., A. Liss, Inc. 1987). In addition, linker-scanning and PCR-mediated techniques can be employed for mutagenesis. See PCR TECHNOLOGY (Erlich Ed., Stockton Press 1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, vols. 1 & 2, *supra*. Protein sequencing, structure and modeling approaches for use with any of the above techniques are disclosed in PROTEIN ENGINEERING, *loc. cit.*, and CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, vols. 1 & 2, *supra*. One preferred variant in this context is a hrZP3 424 amino acid sequence having six histidine amino acids added to the amino terminal end. The added portion allows more convenient affinity purification of the synthesized protein with Ni-NTA affinity resin.

Recombinant hZP3 protein made by cultured cells according to the present invention can be purified by a number of techniques known to the skilled artisan. A preferred method is to collect cell culture media from cells that make the rhZP3 protein and then purify this protein by affinity chromatography. A lectin column, particularly a wheat germ agglutinin column, is preferred for affinity chromatography. Also preferred is to use Ni-NTA affinity resin and to elute protein from this resin at a low pH, in combination with a rhZP3 analog such as that mentioned above having a poly-histidine portion.

The present disclosure permits large-scale expression of biologically active hZP3 glycopolypeptide by recombinant DNA methods. The glycopolypeptide thereby can be obtained in an isolated form by known recombinant methods. The term "isolated" in the context of proteins denotes a degree of purification such that the hZP3 is free of other human proteins that are found with hZP3 in its native context. The isolated protein preferably would be in homogeneous form, that is, in a form amenable to protein sequencing on a gas-phase sequenator, which are available from manufacturers such as Applied Biosystems, Inc. Techniques for obtaining such homogeneity after recombinant production include SDS-PAGE, isoelectric focusing, chromatographic electrophoresis, ion exchange chromatography, gel exclusion chromatography, affinity chromatography, immunoprecipitation, and combinations thereof.

Purify denotes a degree of separation that is higher than isolation. A "purified" protein is sufficiently free of other materials such that the any impurities do not unduly affect the biological properties of the protein or cause other adverse consequences.

The pure or partially purified protein of the invention can be used for many research and clinical purposes. One preferred use is for fertility testing. In this case, rhZP3 is combined with a sample of spermatozoa. Recombinant hZP3 then binds to spermatozoa to produce an acrosome reaction and the reaction is detected. A detectable difference in the proportion of spermatozoa that react indicates male fertility or infertility. Spermatozoa from a fertile male will successfully undergo acrosome reactions at a higher rate than spermatozoa from an infertile male.

In a preferred procedure, semen from a patient donor and control semen from a donor known to have properly functioning sperm, are allowed to liquefy at room temperature. Motile spermatozoa within each liquid sample are selected by mixing the liquid with buffered protein solution followed by centrifugation, an incubation period of a few minutes to allow spermatozoa to swim up, and removal of an upper fluid portion. Active sperm preferably are exposed to 1-100 ng/ml rhZP3 and more preferably to 5-10 ng/ml rhZP3 for at least 5 minutes. After this time, the acrosomal status is determined by comparing control sperm with the patient test sperm by any of a number of techniques known to the skilled artisan.

Both binding and acrosome reaction can be tested for detecting rhZP3 biological activity. For binding tests, rhZP3 is used in detecting the initial stage of binding sperm to zona pellucida. In other embodiments, the recognition and binding of a sperm sample to hrZP3 is directly tested either in solution, or after immobilization of hrZP3 to a solid support such as Sephadex (TM) or an agarose resin. These functional tests evaluate the binding capacity of sperm, and can use (a) rhZP3 conjugated beads (solid phase) or, (b) as part of an ELISA-like test, rhZP3 free in solution (liquid phase). In the latter case, rhZP3 preferably is conjugated to another moiety that can form a signal in the assay. Sperm from infertile male individuals, and that lack ZP3 binding activity will not bind to rhZP3 and hence, can be diagnosed and identified.

Acrosome reaction tests typically link one of the male infertility factors with the inability of sperm to undergo the acrosome reaction, an important phase that contributes to the penetration of sperm into the egg. A preferred embodiment in this context is the

differential staining and visual counting of total cells and acrosome depleted cells after treatment with hrZP3. The etiology of a patient can be determined if the patient's sperm responds to treatment with rhZP3. This response can be detected, for example, by immunofluorescence techniques that are known to the skilled artisan.

In another embodiment, the present invention relates to transformed, human ovarian cell lines, the cells of which contain DNA encoding transfected ZP3. Suitable cell types include but are not limited to, cells of the following types: EB<sub>2</sub>, (human ovary cells, ATCC), CaoV-3, CaoV-4, OVCAR-3, SK-OV-3, SW 626 (human ovary, carcinoma, ATCC). Such cells are described, for example, in the Cell Line Catalog of the American Type Culture Collection (ATCC). The transfer of genes into mammalian cells, to produce a "transfected" cell line, has been well described in the art. See, for example, Ausubel *et al.*, *Introduction of DNA Into Mammalian Cells*, in CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, sections 9.5.1-9.5.6 (John Wiley & Sons, Inc. 1995).

In yet another embodiment, the invention provides therapeutic uses of rhZP3 as a "priming sperm stimulant" in intrauterine insemination therapy ("IUI"). In this procedure, a swim-up procedure is employed to prepare sperm. The inventors have discovered that incubating these sperm with rhZP3 before this swim-up procedure will stimulate an acrosome reaction in some of the sperm and thereby exerts a priming effect on the sperm sample. This priming effect allows the treated sperm to respond more efficiently to motility and acrosome reaction regulators within the female reproductive tract.

In this therapeutic embodiment, sperm are incubated with a rhZP3 dose previously shown to produce a moderate stimulation of the acrosome reaction. Moderate stimulation in this context means that the reaction is stimulated in approximately 10-15 percent of the sperm within a given sample. The incubation step preferably is for at least 30 minutes and is followed by a washing step and the swim-up procedure.

In another therapeutic application, rhZP3 is used to stimulate the acrosome reaction prior to intracytoplasmic sperm injection (ICSI) therapy. Use of rhZP3 in this case can benefit fertilization by, for example, increasing the rate of successful fertilization. Preferably, rhZP3 is added to a suspension of sperm and then washed from this suspension prior to bringing a spermatozoan into an oocyte. In this case, the rhZP3 is used in a high concentration to ensure that a maximum proportion of treated sperm experience an acrosome reaction.

The present invention is further described by reference to the following, illustrative examples, which do not limit the scope of the claimed invention.

**Example 1: Reverse transcriptase polymerase chain reaction of hZP3 cDNA**

A pair of primers (A and B primers) was designed from the reported sequence of hZP3. Chamberlin and Dean *supra*. Primer A is located at the 5' end of the hZP3 cDNA from bases 8 to 29 as 5'ACCATGGAGCTGAGCTATAGG3'. Primer B is located at the 3' end of the hZP3 cDNA from bases 1256 to 1282 as 5'TTCTCGAGTTAACATGATGATGATGATGATGATCGGAAGCAGACACAGGGTG GGAGGCAGT3'. A sequence at an Xhol restriction site (CTCGAG) and a sequence coding for 6 histidine residues were added to the 3' end of primer B for the purpose of subcloning the cDNA into the expression vector as well as for purification of the hZP3 from the medium.

A total RNA extract was isolated from human ovary tissue. RT-PCR was performed and RT-PCR was used as the "standard method" described by Perkin-Elmer Cetus (Foster City, CA). A 1315 base pair DNA fragment comprising the full length of hZP3 cDNA with a 6 histidine tail and an Xhol restriction site was obtained by PCR amplification of the first strand cDNA from the ovary mRNA.

**Example 2: Subcloning hZP3 cDNA into pcDNA 3.1(+) Expression Vector**

The RT-PCR product from Example 1 was separated by agarose gel electrophoresis and purified with a Geneclean II kit (BLO 101, Vista, CA) after cutting out the 1315 bp DNA product. The purified RT-PCR DNA fragment was inserted into a mammalian Expression Vector, pcDNA 3.1(+) (Invitrogen, San Diego, CA) which can express high levels of recombinant protein in mammalian cells and also contains a neomycin resistant gene for selection. The pcDNA 3.1(+) vector with hZP3 cDNA was transformed into E. Coli cells. The positive clones were identified by restriction mapping, southern blot analysis and DNA sequencing.

DNA sequence analysis of hZP3 cDNA revealed that the hZP3 cDNA sequence is identical to that published by Chamberlin and Dean, *Proc. Nat. Acad. Sci. U.S.A.* 87: 6014-6018 (1990). To determine whether the hZP3 cDNA could be translated into a full length recombinant hZP3, in vitro translation was carried out (Promega, Madison, WI). SDS PAGE Analysis of products from the in vitro translation revealed that hZP3 cDNA

produced only a 48kb protein. This 47kb was determined to represent the full length form of recombinant hZP3.

#### Example 3: Stable introduction of rhZP3 into PA-1 cells

The hZP3/pcDNA3.1(+) construct was introduced into cells of an ovarian cell line "PA-1" by the calcium phosphate method. Selection of cells having a stably integrated exogenous gene was carried out with 200uM neomycin (Sigma, St. Louis, MO). The integration of hZP3 into genomic DNA of the host cell was identified by PCR and southern blotting. In the PCR analysis, integration of the hZP3/pcDNA3.1(+) construct was identified using a specific primer which is located on the pcDNA3.1(+) vector (T7 primer, Invitrogen) with primer B, which is located at the 3' end of hZP3. An expected product was obtained only in stably transfected cells but not in untransfected PA-1 cells, indicating successful integration into the PA-1 cell chromosome.

Expression of recombinant hZP3 was detected by RT-PCR and western blot analysis. RT-PCR analysis of mRNA from stably transfected PA-1 cells and non-transfected PA-1 cells with primer A and B revealed that a PCR amplification product was only observed in the stably transfected cells. Culture medium was harvested from both groups of cells, concentrated by speed vacuum and analyzed by SDS PAGE. The samples were blotted to the nitrocellulose and hybridized to a rabbit polyclonal antibody (Ab5a) which can recognize the conserved region of hZP3. A hybridized signal was only observed from transfected PA-1 cells and the solubilized control protein, indicating rhZP3 expression from the stably transformed cells. Cells were stored in liquid nitrogen.

#### Experiment 4: Cell culture of stably transfected PA-1 cells

A stored 1 ml sample of cells was removed from liquid nitrogen and thawed in a 37° water bath for two minutes. The cells were suspended in 5 ml of MEM with 10% FBS and centrifuged at 1000 x g for 5 minutes. The cell pellet was resuspended in 10 ml of MEM with 10% FBS and 200 uM neomycin, and then cultured in a 100mm culture plate. After reaching 90% confluence, cells were removed from the plate using trypsin-EDTA (Sigma, St. Louis, MO) and washed with serum free MEM medium. The cell pellet was resuspended in 60 ml of MEM with 10% FBS and 200 uM neomycin and cultured in three 150 mm culture plates. After reaching 50% confluence, the medium was switched from MEM/10% FBS to protein-free hybridoma medium. The medium was collected every 24

hours and proteinase inhibitors (EDTA, leupeptin, pepstatin, PMSF) were added to protect proteins from proteolytic digestion. The collected media were stored at -20° C until used.

#### Experiment 5: Purification of recombinant hZP3 by affinity chromatography

The collected medium was thawed at 37° C and glycopolypeptides were purified by passage through a WGA affinity column (2 X 4 cm). After applying the medium through the column twice at a flow rate of 3 column volumes per hour, the WGA column was washed with 10mM PBS, 0.15M NaCl, pH 7.2-7.5 until the 280nm absorbance of the effluent fell below 0.01. Glycopolypeptides were eluted with 40mM phosphate and 0.15M NaCl at pH 3.0-3.1 that contained 0.5M N-acetyl-D-glucosamine.

The partially purified protein from the WGA affinity chromatography was dialyzed with 50mM Na-phosphate, 10mM imidazole, and 300mM NaCl overnight at pH 8.0 at 4° C and then loaded onto a Ni-NTA column (Qiagen, Valencia, CA) at 3-4 column volumes per hour. The column was washed with 50mM Na-Phosphate, 300mM NaCl, 10% glycerol at pH 7.8-8.0 until the eluant absorbance at 280 nm fell below 0.01. The hZP3 then was eluted with 200mM Na-phosphate, 300mM NaCl at pH 6.6. The eluted protein was dialyzed with 10mM Na-phosphate pH 7.5 overnight at 4°C. The dialyzed protein was studied by SDS PAGE western blot with an hZP3 antibody. Bands were observed from 65kd to 100kd, indicating the presence of hZP3 isoforms.

#### Experiment 6: Purification of rhZP3 to 80% purity

Agarose-wheat germ agglutinin (WGA) (Vectoe Laboratories, Inc. Burlingame, CA) chromatography was used as a first step to isolate glycoproteins from the cultured cell media. Agarose-wheat germ agglutinin was equilibrated with 10 resin volume of WGA binding buffer (10mM PBS, pH 7.4, 0.15 M NaCl). The collected supernatants were passed through the WGA resin at a flow rate of 3 resin-volumes per hour in a 4°C cold room. The resin was washed with WGA binding buffer until the flow-through 280nm absorbance was less than 0.01. Glycoproteins binding to the WGA resin were separated into two elution peaks with elution buffer A (10 mM PBS, pH 7.4, 0.15 M NaCL, 500 mM-N-acetyl-D-glucosamine (Sigma)) and elution buffer B (10 mM PBS, pH 7.4, 0.15 M

NaCl, 500 mM N-acetyl-D-glucosamine). After eluting glycoproteins from the resin, the resin was washed with WGA binding buffer until the flow-through 280nm absorbance was less than 0.01, and then followed by washing with 5 resin volume of WGA storage buffer [10 mM PBS, pH 7.4, 0.15 M NaCl, 20 mM N-acetyl-D-glucosamine, 0.08% sodium azide (Sigma)]. Eluted glycoprotein samples were dialyzed against the Ni-NTA binding buffer (50 mM PBS, pH 8.0; 300 mM NaCl) for Ni-NTA affinity purification.

Histidine tagged glycoprotein (rhZP3) was purified from the glycoprotein fraction, and isolated from cultured cell media with WGA affinity chromatography, with Ni-NTA (nitrilo-tri-acetic acid) resin (Qiagen). Proteins containing one or more 6xHis affintiy tags, located at either the amino or carboxyl terminates of the protein, bind to the Ni-NTA resin with an affinity ( $K_d = 10 - 13$  at pH 8.0) far greater than the affinity between most antibodies and antigens, or enzymes substrates uspended 50% slurry of Ni-NTA resin were transferred into the column. Two milliliter of resin were completely resuspended then washed with 5 resin-volume of H<sub>2</sub>O. The resin was equilibrated with 10 resin-volumes of Ni-NTA binding buffer. The WGA isolated glycoprotein samples that had been dialyzed with Ni-NTA binding buffer were passed through an equilibrated Ni-NTA column that contained 1 ml of resin. The flow rate was adjusted to 3-4 resin volumes per hour. After passing the glycoprotein sample through the Ni-NTA column, the resin was washed with 10 resin columes of Ni-NTA binding buffer that contained Tween 20 (Fisher) and 2-mercaptopethanol, was followed by a wash with Ni-NTA washing buffer (50 mM PBS, pH 6.6, 300 mM NaCL) until the flow-through 280nm absorbance was less than 0.01. His-tagged glycoproteins were eluted from the resin with Ni-NTA washing buffer that contained 40 mM of imidazole (Sigma).

The centricon 50 filtration was applied to remove co-purified proteins having molecular weights less than 50 kD from the Ni-NTA purified glycoprotein samples. The rhZP3 glycoprotein was found to be 80% pure, as judged by densitometry scanning of an electrophoresis gel of coomassie blue stained protein.

#### Experiment 7: Hemizona Assay ("HZA") of Sperm Biological Activity

Preparation of oocytes. Human immature (prophase I) oocytes were stored in a hyperosmotic solution containing 1.5M MgCl<sub>2</sub> supplemented with 40mM Hepes buffer at pH 7.3 and 0.1% polyvinylpyrrolidone. Oocytes could be stored for up to 90 days at 4°C without affecting performance of the HZA assay. Prior to cutting, oocytes were washed in a culture medium of Ham's F-10 (Gibco Laboratories, Grand Island, N.Y.). Narishige micromanipulators (Tokyo, Japan) mounted on a phase-contrast inverted microscope (Nikon Diaphot, Garden City, N.Y.) were utilized to cut each oocyte into two halves, termed hemizonae.

Preparation of sperm and exposure to 30 ng/ml rhZP3. An aliquot of semen was washed with 2 volumes of Ham's F-10 supplemented with 5% human serum albumin ("HSA"). The sperm suspension was centrifuged for 8 minutes at 400 x g. This wash was repeated. The final pellet was overlaid with 500 ul of Ham's F-10/5% HSA and incubated for one hour in the presence of 5% CO<sub>2</sub> at 37°C. The HSA sperm suspension then was removed and divided into two portions. One portion was used as control sperm (4 hours at 5% CO<sub>2</sub>, 37°C). The second portion was incubated under identical conditions but in the presence of 30ng/ml rhZP3. Supernatants containing motile spermatozoa were removed and a 100 ul droplet of each was separately placed into a petri dish submerged in mineral oil.

Incubation of sperm with hemizonae. One hemizona was transferred to the control droplet and its corresponding hemizona was placed into the rhZP3 treated droplet. Gametes were incubated for 4 hours at 5% CO<sub>2</sub>, 37°C. The hemizonae then were rinsed in culture medium, using a finely drawn glass pipette to dislodge loosely attached sperm.

Results. The number of spermatozoa tightly bound to the convex surface of each hemizona was visually determined with a phase-contrast microscope at 200x magnification. For each sample, a "hemizona index" was calculated as the percentile ratio of the number

of rhZP3-treated sperm bound to the number of non-treated control sperm bound. Defective sperm from infertile patients will generate a lower hemizona index than normal sperm and can be clinically detected this way.

#### Experiment 8: Immunofluorescence assay of acrosome reaction

Semen was obtained from normal donors and allowed to liquefy for 30-60 minutes at room temperature. Semen was divided into 0.5 ml aliquots that were placed at the bottom of plastic 15 ml tubes. Each aliquot was washed with 2 volumes of Ham's F-10 (Sigma) supplemented with 5% human serum albumin. The sperm suspension was centrifuged for 8 minutes at 400 x g. This wash was repeated once. The final pellet was overlaid with 500 ml of Ham's F-10 with 5% human serum albumin. The tubes are loosely capped and incubated at a 30° angle for one hour at 37° C in 5% carbon dioxide, allowing sperm to swim from semen into the medium.

A probe of fluorescein isothiocyanate-conjugated *Pisum Sativum* agglutinin (PSA, Vector Lab, Burlingame, CA) was used to evaluate the acrosomal status of spermatozoa in spot slides. The same slides were counterstained with Hoescht stain, a DNA-specific stain that enters the nuclear membrane of dead spermatozoa, giving a fluorescent counterstain. An epifluorescent microscope was used to read the spot slides at a power of 400 X magnification. Triplicate slides were made for the assay. At least 100 cells were evaluated per spot slide within the grid of the eyepiece of the microscope from 10 random fields. Two trained researchers were assigned to read the results, which were averaged. The results were expressed as percentiles of acrosome-reacted spermatozoa in the total population counted. Spermatozoa that had been treated with cell culture medium from Experiment 4 that contained hZP3 formed significantly more acrosome reactions compared to spermatozoa that had been treatment with a control (no hZP3 protein) medium.

All of the publications (including internet web site) and issued patents cited herein are explicitly incorporated in their entireties by reference.

WE CLAIM:

1. A purified recombinant glycopolypeptide of about 65kd to about 100kd that comprises approximately 40% to 60% carbohydrate by weight and that can strongly bind human spermatozoa and induce an acrosome reaction in the spermatozoa.
2. The glycoprotein of claim 1, wherein the glycoprotein is produced by a human ovarian cell line.
3. A process for producing a glycopolypeptide having the biological activity of hZP3 protein, comprising the steps of:
  - (a) transducing a cell from a human ovarian cell line with a polynucleotide that encodes polypeptide comprising SEQ ID NO:1;
  - (b) establishing a stable-transfected cell culture for producing the glycopolypeptide; and
  - (c) isolating the glycopolypeptide from the cell culture.
4. The process of claim 3, wherein the ovarian cell line is PA-1.
5. The process of claim 3, wherein the polynucleotide encodes a polypeptide comprising the hZP3 sequence.
6. A glycopolypeptide having between 41 and 400 amino acids and having an active portion that preferentially binds to human sperm, wherein the active portion comprises an amino acid sequence that is more than 54% homologous with SEQ ID NO: 1 and has a predicted O-glycosylation site at a serine five positions from the carboxyl terminus of the active portion.
7. The glycopolypeptide of claim 6, wherein the amino acid sequence of the active portion is more than 75% identical with SEQ ID NO: 1.
8. The glycopolypeptide of claim 6, having between 41 and 300 amino acids.
9. The glycopolypeptide of claim 6, having between 41 and 200 amino acids.
10. The glycopolypeptide of claim 6, having between 41 and 100 amino acids.

11. The glycopolypeptide of claim 6, having between 41 and 65 amino acids.

12. A method of detecting male infertility from a human spermatozoa sample, comprising the steps of:

- (a) contacting purified recombinant glycopolypeptide of about 65kd to about 100kd that comprises approximately 50% carbohydrate and that can strongly bind to human spermatozoa and induce an acrosome reaction, with spermatozoa from the sample to form an admixture;
- (b) detecting biological activity of the spermatozoa in the admixture; and
- (c) comparing the biological activity with a reference value.

13. The method of claim 12, wherein step (b) further comprises adding *Pisum Sativum* agglutinin to the admixture.

14. The method of claim 12, wherein the recombinant glycopolypeptide is immobilized onto a solid phase.

15. The method of claim 12, wherein the recombinant glycopolypeptide is made by expression in a human ovarian cell line.

16. The method of claim 12, wherein the human ovarian cell line is PA-1.

17. A nucleic acid vector useful for producing a glycopolypeptide smaller than 300 amino acids long, wherein the glycopolypeptide specifically binds to human sperm and the vector codes for an active amino acid sequence portion that is more than 54% identical with SEQ ID NO: 1.

18. The vector of claim 17, wherein the active amino acid sequence portion is more than 75% identical with SEQ ID NO: 1.

19. The vector of claim 17, wherein the active amino acid sequence portion is 100% identical with SEQ ID NO: 1.

20. A human ovarian cell containing a nucleic acid vector useful for producing a glycopolypeptide less than 300 amino acids long, wherein the glycopolypeptide strongly

binds to human sperm and the vector codes for an amino acid sequence that has a portion which is more than 75% identical with SEQ ID NO: 1.

21. The human cell of claim 20, wherein the amino acid sequence portion is identical with SEQ ID NO: 1.

Table 1

Position No.	Preferred Amino Acid	Other Representative Amino Acids
310	phe	tyr
315	pro	
320	gln	asp
323	asn	gln
324	lys	
326	asp	asn, glu, gln, ile, pro, phe, cys
328	gly	ala, ile
329	thr	ser
330	pro	
332	his	lys
334	arg	lys
335	arg	lys
337	pro	met
339	val	iso, met
341	ser	thr
342	gln	asn
345	arg	lys

Figure 1

Ser Trp Phe Pro Val Gln Gly Pro Ala Asp Ile Cys Gln Cys Cys Asn Lys Gly Asp Gly Thr  
308 318 328

Pro Ser His Ser Arg Arg Glu Pro His Val Met Ser Gln Trp Ser Arg Ser Val Ser

338 . . . . . 348

*Figure 2*

Name: HUMAN Length: 41  
SWFPVQGPADICQCCNKGDCGTPSHSRRQPHVMSQWSRSVS  
.....

Name	Residue No.	Potential	Threshold	Assignment
HUMAN	Thr 22	0.0285	0.4982	

Name	Residue No.	Potential	Threshold	Assignment
HUMAN	Ser 1	0.0629	0.5132	
HUMAN	Ser 24	0.0104	0.5253	
HUMAN	Ser 26	0.0265	0.5309	
HUMAN	Ser 34	0.0033	0.6267	
HUMAN	Ser 37	0.4498	0.5825	
HUMAN	Ser 39	0.0009	0.5126	
HUMAN	Ser 41	0.0082	0.5022	

Figure 3

Name: MOUSE Length: 41  
SWLPVQGDADICDCCSHGNCSN~~SSQFQI~~NGPRQWSKLVS  
.....

Name	Residue No.	Potential	Threshold Assignment
------	-------------	-----------	----------------------

Name	Residue No.	Potential	Threshold Assignment
MOUSE	Ser 1	0.0116	0.5078
MOUSE	Ser 16	0.0002	0.5747
MOUSE	Ser 21	0.0403	0.5845
MOUSE	Ser 23	0.0074	0.5310
MOUSE	Ser 24	0.1625	0.5397
MOUSE	Ser 25	0.0167	0.5585
MOUSE	Ser 26	0.0003	0.5656
MOUSE	Ser 37	0.0024	0.5627
MOUSE	Ser 41	0.0005	0.5127

# INTERNATIONAL SEARCH REPORT

Inte: onal Application No  
PCT/US 99/03273

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC 6 C12N15/12 C07K14/705 G01N33/50 C12N5/10

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 IPC 6 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 92 03548 A (AKZO NV) 5 March 1992 (1992-03-05) page 3, line 1 - line 8 page 7, line 7 - line 10; claims ---	1, 3
A	KINLOCH, R A ET AL.: "Mapping the mouse ZP3 combining site for sperm by exon swapping and site-directed mutagenesis" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 92, 1995, pages 263-267, XP002108572 NATIONAL ACADEMY OF SCIENCE. WASHINGTON., US ISSN: 0027-8424 the whole document ---	1, 4 -/-

Further documents are listed in the continuation of box C.

Patent family members are listed in annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
8 July 1999	21/07/1999
Name and mailing address of the ISA  European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Chambonnet, F

## INTERNATIONAL SEARCH REPORT

Inte	ional Application No
PCT/US 99/03273	

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	CHAMBERLIN, M. E. & DEAN, J.: "Human homolog of the mouse sperm receptor" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA., vol. 87, August 1990 (1990-08), pages 6014-6018, XP002108573 NATIONAL ACADEMY OF SCIENCE. WASHINGTON., US ISSN: 0027-8424 cited in the application the whole document ----	1,3
A	VAN DUIN, M. ET AL.: "Cloning and characterization of the human sperm receptor ligand ZP3: evidence of a second polymorphic allele with different frequency in the caucasian and the japanese populations" GENOMICS, vol. 14, no. 4, December 1992 (1992-12), pages 1064-1, XP002108574 the whole document ----	1
A	BAGAVANT, H. ET AL.: "Immunogenicity and contraceptive potential of a zona Pellucida 3 peptide vaccine" BIOLOGY OF REPRODUCTION, vol. 56, no. 3, March 1997 (1997-03), pages 764-770, XP002108575 the whole document -----	1

# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 99/03273

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO 9203548	A 05-03-1992	AU	8328591 A	17-03-1992
		CA	2090486 A	28-02-1992
		CN	1060499 A	22-04-1992
		FI	930897 A	05-04-1993
		HU	64394 A	28-12-1993
		JP	6500690 T	27-01-1994
		NZ	239518 A	25-03-1994
		PT	98780 A	31-07-1992